Operator Splitting Methods in Control
Operator Splitting Methods in Control

Giorgos Stathopoulos
École Polytechnique Fédérale de Lausanne (EPFL)

Harsh Shukla
École Polytechnique Fédérale de Lausanne (EPFL)

Alexander Szűcs
Slovak University of Technology in Bratislava

Ye Pu
École Polytechnique Fédérale de Lausanne (EPFL)

Colin N. Jones
École Polytechnique Fédérale de Lausanne (EPFL)
Editorial Scope

Topics

Foundations and Trends® in Systems and Control publishes survey and tutorial articles in the following topics:

- Control of:
  - Hybrid and discrete event systems
  - Nonlinear systems
  - Network systems
  - Stochastic systems
  - Multi-agent systems
  - Distributed parameter systems
  - Delay systems
- Systems
  - Energy storage
  - Grid integration
  - Conversion technologies
  - Underpinning materials developments

Filtering, estimation, and identification
- Optimal control
- Systems theory
- Control applications

Information for Librarians

Foundations and Trends® in Systems and Control, 2016, Volume 3, 4 issues. ISSN paper version 2325-6818. ISSN online version 2325-6826. Also available as a combined paper and online subscription.
Operator Splitting Methods in Control

Giorgos Stathopoulos
École Polytechnique Fédérale de Lausanne (EPFL)

Harsh Shukla
École Polytechnique Fédérale de Lausanne (EPFL)

Alexander Szűcs
Slovak University of Technology in Bratislava

Ye Pu
École Polytechnique Fédérale de Lausanne (EPFL)

Colin N. Jones
École Polytechnique Fédérale de Lausanne (EPFL)
# Contents

1 Introduction
   1.1 Notation and Definitions ................................... 7

2 The Algorithms
   2.1 The dual problem ............................................. 10
   2.2 Proximal methods ............................................. 12
   2.3 Origin of the methods and a unified framework ............. 14
   2.4 Alternative interpretations .................................. 18
   2.5 Relaxation .................................................. 20
   2.6 Termination ................................................ 20
   2.7 Discussion ................................................ 21

3 Convergence Results and Accelerated Variants ................. 25
   3.1 Sublinear convergence ....................................... 27
   3.2 Accelerated sublinear convergence .......................... 28
   3.3 Linear convergence ......................................... 35
   3.4 Discussion ................................................ 36

4 Stepsize Selection and Preconditioning ........................ 39
   4.1 Stepsize .................................................. 39
   4.2 Preconditioning ............................................ 43
   4.3 General functions .......................................... 50
Abstract

The significant progress that has been made in recent years both in hardware implementations and in numerical computing has rendered real-time optimization-based control a viable option when it comes to advanced industrial applications. More recently, the need for control of a process in the presence of a limited amount of hardware resources has triggered research in the direction of embedded optimization-based control. At the same time, and standing at the other side of the spectrum, the field of big data has emerged, seeking for solutions to problems that classical optimization algorithms are incapable to provide. This triggered some interest to revisit the family of first order methods commonly known as decomposition schemes or operator splitting methods. Although it is established that splitting methods are quite beneficial when applied to large-scale problems, their potential in solving small to medium scale embedded optimization problems has not been studied so extensively. Our purpose is to study the behavior of such algorithms as solvers of control-related problems of that scale. Our effort focuses on identifying special characteristics of these problems and how they can be exploited by some popular splitting methods.

DOI: 10.1561/2600000008.
The significant progress that has been made in recent years both in hardware implementations and in numerical computing has rendered real-time optimization-based control a viable option when it comes to advanced industrial applications. More recently, the need for control of a process in the presence of a limited amount of hardware resources has triggered research in the direction of embedded optimization-based control. Many efficient high-speed solvers have been developed for both linear and nonlinear control, based on either first order methods (FiOrdOs [133], QPgen [57, 59], DuQuad [88]), interior point (IP) methods (FORCES [43], CVXGEN [84]) and active set methods (QPOASES [50]).

In this work we focus on systems with linear dynamics, giving rise to convex control problems. The purpose of the survey is to explore a family of first order methods known as decomposition schemes or operator splitting methods. The abstract form of the problem at hand is the minimization of the sum of two convex functions subject to linear equality constraints, and can be written as

$$\text{minimize} \quad f(z) + g(Lz) \ ,$$

(1.1)
with variables $z \in \mathbb{R}^n$, where $f$ and $g$ are closed, proper convex functions and $A : \mathbb{R}^n \to \mathbb{R}^p$ is a linear map. A splitting method can be applied to the above problem after rewriting it as

$$\begin{align*}
\text{minimize} & \quad f(z) + g(y) \\
\text{subject to} & \quad Lz = y ,
\end{align*}$$

by alternatingly (or simultaneously) minimizing over $y$ and $z$. Clearly, the solutions of problems (1.2) and (1.4) are identical. Inequality constraints that might appear are already embedded in one of the two functions in the form of indicator functions, i.e., a membership function for a set $C$

$$\delta_C(z) = \begin{cases} 
0 & z \in C \\
\infty & \text{otherwise},
\end{cases}$$

which is the reason why both $f$ and $g$ are considered to be extended-real-valued functions (see [18, § 3.1.2]). Formulations similar to the above have been studied extensively and we can look for their roots in the method of multipliers [75], [110], the Arrow-Hurwicz method [9], Douglas-Rachford splitting [44] and ADMM [60], [55]. Decomposition of the original problem into simpler ones is beneficial when distributed computation tools are available. This potential is already suggested in the classical references [15] and [15]. It was not until recently, though, that decomposition algorithms were indeed applied in modern engineering problems (signal and image processing, big data analysis, machine learning, [17] and [27]), in cases where off-the-shelf interior point solvers simply fail due to the large dimensions involved. The thesis [47] provides a comprehensive description of the connection of several splitting algorithms under a common framework. Finally, the book [7] provides a mathematically rigorous introduction to operator splitting methods in general Hilbert spaces.

The plethora of different approaches for solving problem (1.2) is partly a consequence of the problem-dependent behavior of first order methods. This behavior has both its pros and cons; on one hand, sensitivity to the problem’s structure and data requires pre-processing and tuning of several parameters, a procedure that can be cumbersome. However, it is exactly this procedure that gives the flexibility to customize the solver to the problem at hand, and, in many cases,
outperform by several orders of magnitude general purpose solvers. Consequently, there are numerous approaches, each of which can be less or more pertinent for the specific problem. Mentioning some of the most important categorizations, we can solve either the primal problem, the dual problem, or a primal-dual formulation. Regarding primal approaches, the most popular one is the primal decomposition method [15], [19], where the original problem is decomposed into a master problem and two subproblems. The two subproblems have both local and shared (complicating) variables, while the master subproblem manipulates only the complicating variables. Primal decomposition works well when the complicating variables for the two subproblems are few.

Dualization plays a crucial role in more complicated problems. It can be performed by means of Lagrangian relaxations (dual decomposition [35], [49], [123], [14]), augmented Lagrangian relaxations [13], [117], [116], alternating minimization (Gauss-Seidel) augmented Lagrangian schemes (ADMM), mixture of Lagrangian with augmented Lagrangian schemes (AMA [131]), linearized augmented Lagrangians or approximate minimization schemes (23, 4) and, finally, mixtures of alternating minimization with partial linearization (PDHG [139], [48], [22], [30] and several similar primal-dual schemes [28], [134], [16]).

Although it is well-established that splitting methods are quite beneficial when applied to large-scale problems, their potential in solving small to medium scale embedded optimization problems has not been studied so extensively. It was not until very recently that the first works attempting to apply decomposition methods in control problems started making their appearance [102], [56], [57], [59], [105]. Our purpose is to study the behavior of such algorithms as solvers of control-related convex problems of that scale, i.e., from tens to a few hundreds of variables. Our effort focuses on identifying special characteristics of these problems and how they can be exploited by some popular splitting methods. Some of the questions that we attempt to answer are:

1. It is very common in practice that optimal control problems come with a quadratic objective, since in this way stability can be proven for regulation or tracking purposes. What is the best way
to exploit this smooth term, along with the special structure of the dynamics equation?

2. Given that a control problem has to be solved repeatedly (e.g., MPC), how does warm-starting of the solution affect the speed?

3. Given the structure of the problem at hand, which algorithms will converge more quickly?

4. Are there ways to precondition the problem in order to reduce the solve time?

In what follows we present three well-understood splitting algorithms, the alternating direction method of multipliers (ADMM), the alternating minimization algorithm (AMA) and a primal-dual algorithm (PDA), the most popular representative of several primal-dual schemes that have been recently developed. These three methods come from different sides of the spectrum described above, but also hold very strong similarities. Our choice is motivated from the fact that the methods are analyzed and extended from several communities, and hence their properties are well-understood.

The paper is organized as follows: In Chapter 2 we formulate the problem we want to solve and look at it from three different perspectives, resulting in the three algorithms we use. Subsequently we introduce the algorithms under a unified scheme and report their properties. In Chapters 3 and 4 we build on the basic variants of the methods presented before, introduce several enhanced versions and focus on their applicability for solving optimization problems. More specifically, in Chapter 3 we review how one can exploit the structure of the problem to accelerate the theoretical convergence rates. In Chapter 4 we extend the discussion on acceleration to more practical schemes, i.e., stepsize selection and preconditioning. We provide a comprehensive literature review of existing methods and we present generic preconditioned versions of the three algorithms. In Chapter 5 we discuss the computational aspects; we identify the bottlenecks in each method and propose ways to speed up the computation. In Chapter 6 we summarize the
1.1. Notation and Definitions

Let $\mathcal{Z} = (\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ be a Euclidean space equipped with the inner product $\langle z, x \rangle = z^T x$ and the corresponding norm $\|z\| = \sqrt{\langle z, z \rangle}$. Symmetric $n$-dimensional matrices are denoted with $\mathbb{S}^n$, while positive (semi)definite matrices are denoted with $(\mathbb{S}^n_+) \mathbb{S}^n_{++}$. We also consider the scaled norm $\|z\|_P = \sqrt{\langle z, Pz \rangle}$, with $P \in \mathbb{S}_+$. The matrix norm of the linear operator $M \in \mathbb{R}^{m \times n}$ is defined as $\|M\| = \sup_{z \neq 0} \frac{\|Mz\|}{\|z\|}$. The minimum and maximum eigenvalue of a matrix $Q \in \mathbb{R}^{n \times n}$ are denoted by $\lambda_{\min}(Q)$ and $\lambda_{\max}(Q)$, respectively.

The domain of the extended-real-valued function $f$ is defined as $\text{dom} f = \{z \in \mathcal{Z} : f(z) < +\infty\}$ and $f$ is proper if $\text{dom} f \neq \emptyset$ and $f > -\infty$. The function $f$ is closed if its epigraph $\text{epi} f = \{(z, t) \in \mathbb{R}^n \times \mathbb{R} : f(z) \leq t\}$ is a closed nonempty convex set. The range of extended-real-valued functions is denoted with $\mathbb{R} \cup \{+\infty\} = \bar{\mathbb{R}}$. We denote the conjugate of a convex function with $f^*$, while a minimizer is denoted by an asterisk, i.e., $f(z^*) \leq f(z) \ \forall z \in \mathcal{Z}$. Finally, for succinctness in the notation, we denote the class of all proper, closed, convex functions from $\mathcal{Z}$ to $\bar{\mathbb{R}}$ with $\Gamma_0(\mathcal{Z})$.

The indicator function of a convex set $C$ is denoted with $\delta_C(\cdot)$. For the common norm balls the notation changes to $\delta_i(z, \alpha)$, $i = 1, 2, \infty$, which denotes the constraint $\|z\|_i \leq \alpha$. Similar notation to the 2-norm ball is used for the second-order cone constraint, with the difference that the second argument is a scalar affine function itself, i.e., $\delta_2(Ax + b, c^T z + d)$ denotes the constraint $\|Az + b\|_2 \leq c^T z + d$. The most common pairs of indicator functions with their conjugate representation are given in Table A.1.
References


References


[60] Marroco A. Glowinski, R. Sur l’approximation, par éléments finis
d’ordre un, et la résolution, par pénalisation-dualité d’une classe de
problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Mod-
elling and Numerical Analysis - Modélisation Mathématique et Analyse

[61] R. Glowinski and P. Le Tallec. *Augmented Lagrangian And Operator-
splitting Methods In Nonlinear Mechanics*. Society for Industrial and

brid gradient methods for saddle-point problems. *arXiv preprint

alternating direction optimization methods. *SIAM Journal on Imaging

[64] T. Goldstein and S. Osher. The Split Bregman Method for $l_1$-
343, 2009.

Backward Splitting with a FASTA Implementation. *arXiv preprint


simulation environment for building control. In *The Multi-Conference
on Systems and Control*, 2015.

[68] O. Güler. On the Convergence of the Proximal Point Algorithm for
Convex Minimization. *SIAM Journal on Control and Optimization*,


[70] A. Hamdi and S. K. Mishra. *Decomposition Methods Based on Aug-

[71] E. N. Hartley, J. L. Jerez, A. Suardi, J. M. Maciejowski, E. C. Kerri-
gan, and G. Constantinides. Predictive Control using an FPGA with
Application to Aircraft Control. *IEEE Transactions on Control Systems
Technology*, 2013.
References


References


References

[111] Y. Pu, M. N. Zeilinger, and C. N. Jones. Inexact Fast Alternating Min-
imization Algorithm for Distributed Model Predictive Control. In The
53rd IEEE Annual Conference on Decision and Control, pages 5915–
5921, 2014.

[112] A. U. Raghunathan and S. Di Cairano. Infeasibility detection in alter-
nating direction method of multipliers for convex quadratic programs.
In The 53rd IEEE Annual Conference on Decision and Control, pages

alternating direction method of multipliers for convex quadratic pro-
gramms and model predictive control. In International Symposium on


[116] R. T. Rockafellar. Augmented Lagrangians and Applications of the
Proximal Point Algorithm in Convex Programming. Mathematics of

[117] R. T. Rockafellar. Monotone Operators and the Proximal Point Algo-

with Saddle Functions and Minimax Problems, pages 241–250. Ameri-
can Mathematical Society, 1970.


[120] R. Scattolini. Architectures for distributed and hierarchical model pre-
dictive control - a review. Journal of Process Control, 19(5):723–731,
2009.

[121] M. Schmidt, N. L. Roux, and F. Bach. Convergence rates of inex-
act proximal-gradient methods for convex optimization. In The 25th
Annual Conference on Neural Information Processing Systems, pages
6819–6824, 2011.

[122] R. Shefi and M. Teboulle. Rate of Convergence Analysis of Decomposi-
tion Methods Based on the Proximal Method of Multipliers for Convex

References


References

